Homework 3 Report Zhengxian Lin

Part1

I did all of l three models for planner, and the results are saved at txt files. I will focus on the result of policy iteration.

For running, run the main.py, and the result files will be created. I will submit the result files too. The reason why I saved all result in file is that the results of the part 3 are large.

Part 2

The results of three models of MDP1-hw3.txt and MDP2-hw3.txt are showed at below. The results also can be find at the txt results files I submitted.

MDP1 (beta = 0.1, ε = 10^{-5}):

```
------------------Value Iteration------------------
Value Function:
0.1001 0.0090 0.0088 0.0090 1.0010 0.0068 0.0683 0.0086 0.0100 0.0896
Policy:
3 3 2 0 0 0 1 2 1 3
------------------Policy Iteration------------------
Value Function:
0.1001 0.0090 0.0088 0.0090 1.0010 0.0068 0.0683 0.0086 0.0100 0.0896
Policy:
3 3 2 0 0 0 1 2 1 3
------------------Modified Policy Iteration------------------
Value Function:
0.1001 0.0090 0.0088 0.0090 1.0010 0.0068 0.0683 0.0086 0.0100 0.0896
Policy:
3 3 2 0 0 0 1 2 1 3
```

MDP1 (beta = 0.9, ε = 10^{-5}):

```
------------------Value Iteration------------------
Value Function:
3.3210 2.9234 2.8914 2.9234 3.6900 2.8407 3.1564 2.9071 2.9889 3.2482
Policy:
3 3 2 0 0 0 1 2 1 3
------------------Policy Iteration------------------
Value Function:
3.3210 2.9234 2.8914 2.9234 3.6900 2.8407 3.1564 2.9071 2.9889 3.2482 \
Policy Function:
3 3 2 0 0 0 1 2 1 3
------------------Modified Policy Iteration------------------
Value Function:
3.3210 2.9234 2.8914 2.9234 3.6900 2.8407 3.1564 2.9071 2.9889 3.2482
Policy Function:
3 3 2 0 0 0 1 2 1 3
```

MDP2 (beta = 0.1, ε = 10^{-5}):

```
------------------Value Iteration------------------
Value Function:
0.0114 0.0100 0.5733 0.0015 0.0604 0.1010 1.0101 0.0080 0.0060 0.1010
Policy Function:
3 0 0 1 2 2 2 3 1 2
------------------Policy Iteration------------------
Value Function:
0.0114 0.0100 0.5733 0.0015 0.0604 0.1010 1.0101 0.0080 0.0060 0.1010
Policy Function:
3 0 0 1 2 2 2 3 1 2
------------------Modified Policy Iteration------------------
Value Function:
0.0114 0.0100 0.5733 0.0015 0.0604 0.1010 1.0101 0.0080 0.0060 0.1010
Policy Function:
3 0 0 1 2 2 2 3 1 2
```

MDP2 (beta = 0.9, ε = 10^{-5}):

```
------------------Value Iteration------------------
Value Function:
4.2632 4.2577 5.1885 3.8440 4.4169 4.7368 5.2632 3.8351 3.9752 4.7368
Policy Function:
0 0 0 1 2 2 2 1 1 2
------------------Policy Iteration------------------
Value Function:
4.2632 4.2577 5.1885 3.8440 4.4169 4.7368 5.2632 3.8351 3.9752 4.7368
Policy Function:
0 0 0 1 2 2 2 1 1 2
------------------Modified Policy Iteration------------------
Value Function:
4.2632 4.2577 5.1885 3.8440 4.4169 4.7368 5.2632 3.8351 3.9752 4.7368
Policy Function:
0 0 0 1 2 2 2 1 1 2
```

Part 3

The both MDP I design for n = 10, which means there are 10 spots of each line, A and B. In other word, there are 20 spots. Following the requirement on the pdf, the A[1] and B[1] are handicap, and the possibility of occupied is decreased, and the reward of spots are increased except the A[1] and B[1], while the spot closer to the store.

There are three parameters of my MDP. The first one is n, which equals 10. The second one and third one is parameter set. They are same.

For the parameter set, paras = (p1, p2, p3, p4, p5, p6), the meaning of each sub parameter in the tuple is:

p1: The decreasing value of probability of occupation from the closer to the farther. For example, if

p1 = 0.1, and the probability of occupation of A[2] is 0.9, then and the probability of occupation of A[3] is 0.8 = 0.9 – 0.1.
p2: The value of possibility of Occupation of Handicap for both A[1] and B[1]
p3: A tuple (p3-0, p3-1). p3-0 is a highest value of spots, which is A[2] and B[2]; the p3-1 is the decreasing value of reward of spots. While the spot is further, the reward is less. For example, the reward of A[2] = p3-0 = 100, and the p3-1 is 10. Then, the reward of A[3] is 100 – 10 = 90
p4: Driving cost. The cost of driving action
p5: Collision Cost. The cost of parking action at a spot which already had car
p6: The cost or reward of parking action of Handicap spots, A[1] and B[1]

I created two MDP model using two different parameters set and same n = 10. The beta is 0.999 and ε is 10^{-5}. It considers the future a lot. The result of both of them depend on the policy iteration with using value iteration on policy evaluation.
For the same of both MDP, the probability of occupied by another car are same, the decreasing value are 1 / (n -1) which is 1 / 9. The probability of occupation of Handicap are same, 0.01. The Collision Cost are same, since no one like to hit another car, and the punishment is heavy, -10000. The cost of parking action of Handicap spots are same, -100. It is not moral to park at that spot

1. The first one I named it "hating walking driver". The driver very hate walking, and the reward of closer spot has reward to him. The driving action cost too small, just 0.01.
paras1 = (1 / (n - 1), 0.01, (1000, 1000 / (n - 1)), -0.01, -10000, -100)
Paking_MDP1 (beta = 0.999, ε = 10^{-5}):

```
------------------ Policy Iteration------------------
Value Function:
(A-10, unoccupied, unparked):921.8649
(A-9, unoccupied, unparked):922.7877
(A-8, unoccupied, unparked):923.7114
(A-7, unoccupied, unparked):924.6361
(A-6, unoccupied, unparked):925.5616
(A-5, unoccupied, unparked):926.4881
(A-4, unoccupied, unparked):927.4155
(A-3, unoccupied, unparked):928.3439
(A-2, unoccupied, unparked):1000.0000
(A-1, unoccupied, unparked):921.3547
(B-1, unoccupied, unparked):922.2769
(B-2, unoccupied, unparked):1000.0000
(B-3, unoccupied, unparked):914.5158
(B-4, unoccupied, unparked):915.4312
(B-5, unoccupied, unparked):916.3475
(B-6, unoccupied, unparked):917.2648
(B-7, unoccupied, unparked):918.1830
(B-8, unoccupied, unparked):919.1021
(B-9, unoccupied, unparked):920.0221
(B-10, unoccupied, unparked):920.9431
```

```
(A-10, occupied, unparked):921.8649
(A-9, occupied, unparked):922.7877
(A-8, occupied, unparked):923.7114
(A-7, occupied, unparked):924.6361
(A-6, occupied, unparked):925.5616
(A-5, occupied, unparked):926.4881
(A-4, occupied, unparked):927.4155
(A-3, occupied, unparked):928.3439
(A-2, occupied, unparked):920.4333
(A-1, occupied, unparked):921.3547
(B-1, occupied, unparked):922.2769
(B-2, occupied, unparked):913.6012
(B-3, occupied, unparked):914.5158
(B-4, occupied, unparked):915.4312
(B-5, occupied, unparked):916.3475
(B-6, occupied, unparked):917.2648
(B-7, occupied, unparked):918.1830
(B-8, occupied, unparked):919.1021
(B-9, occupied, unparked):920.0221
(B-10, occupied, unparked):920.9431
terminal(-, -, Parked) : 0.0000
Policy:
(A-10, unoccupied, unparked):Drive
(A-9, unoccupied, unparked):Drive
(A-8, unoccupied, unparked):Drive
(A-7, unoccupied, unparked):Drive
(A-6, unoccupied, unparked):Drive
(A-5, unoccupied, unparked):Drive
(A-4, unoccupied, unparked):Drive
(A-3, unoccupied, unparked):Drive
(A-2, unoccupied, unparked):Park
(A-1, unoccupied, unparked):Drive
(B-1, unoccupied, unparked):Drive
(B-2, unoccupied, unparked):Park
(B-3, unoccupied, unparked):Drive
(B-4, unoccupied, unparked):Drive
(B-5, unoccupied, unparked):Drive
(B-6, unoccupied, unparked):Drive
(B-7, unoccupied, unparked):Drive
(B-8, unoccupied, unparked):Drive
(B-9, unoccupied, unparked):Drive
(B-10, unoccupied, unparked):Drive
(A-10, occupied, unparked):Drive
(A-9, occupied, unparked):Drive
```

```
(A-8, occupied, unparked):Drive
(A-7, occupied, unparked):Drive
(A-6, occupied, unparked):Drive
(A-5, occupied, unparked):Drive
(A-4, occupied, unparked):Drive
(A-3, occupied, unparked):Drive
(A-2, occupied, unparked):Drive
(A-1, occupied, unparked):Drive
(B-1, occupied, unparked):Drive
(B-2, occupied, unparked):Drive
(B-3, occupied, unparked):Drive
(B-4, occupied, unparked):Drive
(B-5, occupied, unparked):Drive
(B-6, occupied, unparked):Drive
(B-7, occupied, unparked):Drive
(B-8, occupied, unparked):Drive
(B-9, occupied, unparked):Drive
(B-10, occupied, unparked):Drive
terminal(-, -, Parked) : None
```

Conclusion: The driver will keeping driving until the for the A[2] and B[2], and keep driving at the handicap and occupied spots. Since it hating walking, so keep driving until the closest normal spot is available, which is A[2], not A[1]. That is reasonable

2. The second driver, I call it "hating driving driver", but I do not why he drive car to store. The closer spot seems not attract the driver so much. So, the reward is not too high, starting from 200, decreasing 200 / (n - 1) each spot. The cost of driving is big, -30 for each step.

```
paras2 = (1 / (n - 1), 0.01, (200, 200 / (n - 1)), -30, -10000, -100)
```

Paking_MDP2 (beta = 0.999, $\varepsilon = 10^{-5}$):

```
------------------ Policy Iteration------------------
Value Function:
(A-10, unoccupied, unparked):22.2222
(A-9, unoccupied, unparked):44.4444
(A-8, unoccupied, unparked):66.6667
(A-7, unoccupied, unparked):88.8889
(A-6, unoccupied, unparked):111.1111
(A-5, unoccupied, unparked):133.3333
(A-4, unoccupied, unparked):155.5556
(A-3, unoccupied, unparked):177.7778
(A-2, unoccupied, unparked):200.0000
(A-1, unoccupied, unparked):11.2945
(B-1, unoccupied, unparked):41.3358
(B-2, unoccupied, unparked):200.0000
(B-3, unoccupied, unparked):177.7778
```

```
(B-4, unoccupied, unparked):155.5556
(B-5, unoccupied, unparked):133.3333
(B-6, unoccupied, unparked):111.1111
(B-7, unoccupied, unparked):88.8889
(B-8, unoccupied, unparked):66.6667
(B-9, unoccupied, unparked):44.4444
(B-10, unoccupied, unparked):22.2222
(A-10, occupied, unparked):13.0037
(A-9, occupied, unparked):31.8643
(A-8, occupied, unparked):45.3324
(A-7, occupied, unparked):48.4415
(A-6, occupied, unparked):37.7739
(A-5, occupied, unparked):15.4579
(A-4, occupied, unparked):-9.5145
(A-3, occupied, unparked):-24.4230
(A-2, occupied, unparked):-18.7168
(A-1, occupied, unparked):11.2945
(B-1, occupied, unparked):41.3358
(B-2, occupied, unparked):55.3349
(B-3, occupied, unparked):59.0359
(B-4, occupied, unparked):55.9147
(B-5, occupied, unparked):48.1414
(B-6, occupied, unparked):37.0979
(B-7, occupied, unparked):23.7108
(B-8, occupied, unparked):8.6014
(B-9, occupied, unparked):-7.8000
(B-10, occupied, unparked):-7.8000
terminal(-, -, Parked) : 0.0000
Policy:
(A-10, unoccupied, unparked):Park
(A-9, unoccupied, unparked):Park
(A-8, unoccupied, unparked):Park
(A-7, unoccupied, unparked):Park
(A-6, unoccupied, unparked):Park
(A-5, unoccupied, unparked):Park
(A-4, unoccupied, unparked):Park
(A-3, unoccupied, unparked):Park
(A-2, unoccupied, unparked):Park
(A-1, unoccupied, unparked):Drive
(B-1, unoccupied, unparked):Drive
(B-2, unoccupied, unparked):Park
(B-3, unoccupied, unparked):Park
(B-4, unoccupied, unparked):Park
(B-5, unoccupied, unparked):Park
```

```
(B-6, unoccupied, unparked):Park
(B-7, unoccupied, unparked):Park
(B-8, unoccupied, unparked):Park
(B-9, unoccupied, unparked):Park
(B-10, unoccupied, unparked):Park
(A-10, occupied, unparked):Drive
(A-9, occupied, unparked):Drive
(A-8, occupied, unparked):Drive
(A-7, occupied, unparked):Drive
(A-6, occupied, unparked):Drive
(A-5, occupied, unparked):Drive
(A-4, occupied, unparked):Drive
(A-3, occupied, unparked):Drive
(A-2, occupied, unparked):Drive
(A-1, occupied, unparked):Drive
(B-1, occupied, unparked):Drive
(B-2, occupied, unparked):Drive
(B-3, occupied, unparked):Drive
(B-4, occupied, unparked):Drive
(B-5, occupied, unparked):Drive
(B-6, occupied, unparked):Drive
(B-7, occupied, unparked):Drive
(B-8, occupied, unparked):Drive
(B-9, occupied, unparked):Drive
(B-10, occupied, unparked):Drive
terminal(-, -, Parked) : None
```

Conclusion: the driver will park at the first spot which is not occupied by another car since he hates driving. However, the cost of handicap result in that he will not park at there, because the punishment of both of them are heavy.

Both of agents will avoid the Collision and Handicap spot. The decisions of two of them are reasonable, depend on their personality, like driving or hating driving.

I think that is interesting to do more research in this problem. In other word, creating more different agents. Some maybe think it is ok to park their car in handicap spot.